DOCUMENT RESUME

ED 081 405 HE 004 625

AUTHOR	Aleamoni, Lawrence M.; Hexner, Pamela Z.
TITLE	The Effect of Different Sets of Instructions on Student Course and Instructor Evaluation.
INSTITUTION	Illinois Univ., Urbana. Office of Instructional Resources.
REPORT NO	RR-339
PUB DATE	May 73
NOTE	16p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Course Evaluation; Evaluation; Evaluation Methods; *Higher Education; *Measurement Techniques; Questionnaires; *Student Opinion; Teacher Evaluation; *Validity

ABSTRACT

Two groups of students registered in the fall and spring semesters, respectively, of an undergraduate history course taught by the same instructor, evaluated the course by means of the Illinois Course Evaluation Questionnaire. Students registered in the spring were informed that the results of the evaluation would be used for consideration of the promotion and advancement of the instructor, while students registered in the fall were not provided this information. The 224 results indicated that students who were given this information rated the course more favorably on all aspects than students lacking this information. The results do suggest that the instructions and the information about the aims of a course evaluation that students are provided prior to the completion of an evaluation are important variables that will have to be considered in future administration of any course evaluation. (Author)

ED 081405

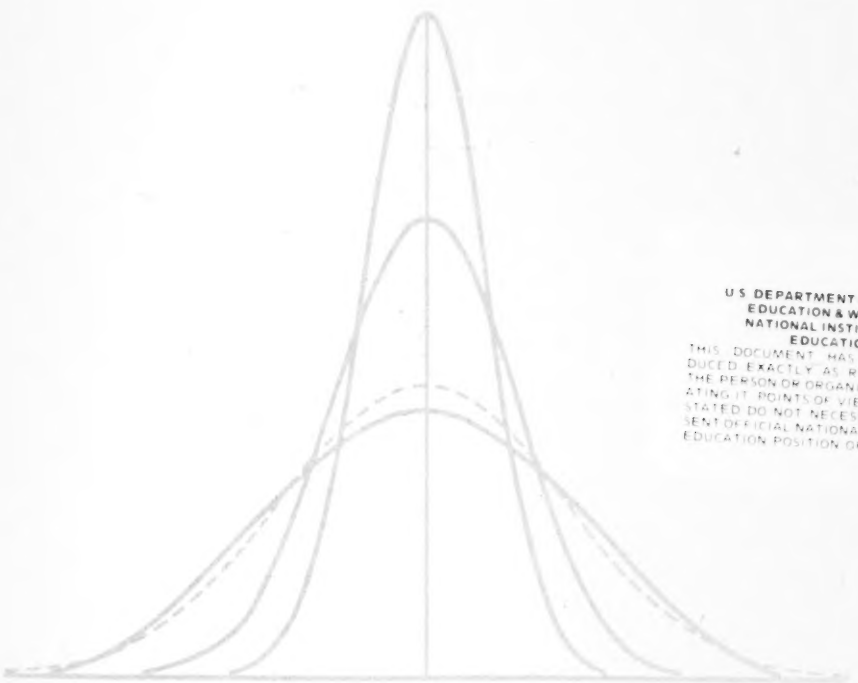
Research Report

TITLE

THE EFFECT OF DIFFERENT SETS OF INSTRUCTIONS ON STUDENT COURSE AND INSTRUCTOR EVALUATION

AUTHOR(S)

Lawrence M. Aleamoni and Pamela Z. Hexner



U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

REPORT # **339**

DATE **May, 1973**

May be quoted in whole or in part if credit is given.

MEASUREMENT AND RESEARCH DIVISION - OFFICE OF INSTRUCTIONAL RESOURCES
307 ENGINEERING HALL - UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS 61801

HE 004 625

FILMED FROM BEST AVAILABLE COPY

Abstract

Two groups of students registered in the fall and spring semesters, respectively, of an undergraduate history course taught by the same instructor, evaluated the course by means of the Illinois Course Evaluation Questionnaire. Students registered in the spring were informed that the results of the evaluation would be used for consideration of the promotion and advancement of the instructor, while students registered in the fall were not provided this information. The results indicated that students who were given this information rated the course more favorably on all aspects than students lacking this information.

THE EFFECT OF DIFFERENT SETS OF INSTRUCTIONS ON STUDENT COURSE AND INSTRUCTOR EVALUATION[1]

Lawrence M. Aleamoni and Pamela Z. Hexner

In the mad rush to make courses "relevant" and meet new demands of accountability, colleges and universities have proposed many methods of evaluating the effectiveness of instruction. Such proposals generally indicate that many elements of the instructional setting need to be evaluated by several different audiences. Unfortunately, most proposals that are operationalized rest solely on the use of student ratings of teachers and informal colleague opinions. That students are able to provide reliable and valid evaluations of instructional quality has come to be recognized (Costin, Greenough, & Menges, 1971; Aleamoni, 1972a).

Much of the research on student rating of teachers has been concerned with the effect of different variables on these ratings. Perhaps the variable of greatest interest has been the grade received or expected by the student raters. Several studies have reported no relationship between students' ratings of instructors and their expected or actual grades (e.g., Bendig, 1953a; Cohen & Humphreys, 1960; Remmers, 1928, 1930, 1939, 1960). Other investigators have found significant positive relationships between students' grades and their evaluation of instructors (e.g., Anikeef, 1953; Spencer & Aleamoni, 1969; Weaver, 1960). Usually, the magnitude of the relationships are small, seldom exceeding .30 (Spencer & Aleamoni, 1969) or ranging from .23 to .32 (Caffrey, 1969).

---

[1] The authors are indebted to Professor Robert A. Waller for cooperating in obtaining the data on his two history courses.

In a recent review of the literature, Costin, Greenough and Menges (1971) have enumerated some variables found to affect instructor evaluation:

1. The rank of the instructor. [Full professors have tended to receive a higher rating than other ranks (Downie, 1952; Gage, 1961).]
2. The level of study completed by the student raters. [Although some studies have not found any differences in evaluation for students in different college years (Guthrie, 1954; Heilman & Armentrout, 1936; Stewart & Melpass, 1966; Rayder, 1968), other investigators (Clark & Keller, 1954; Downie, 1952; Gage, 1961; Lovell & Haner, 1955) have found that more advanced students tended to give more favorable ratings than less advanced students and graduate students have evaluated instructors more favorably than undergraduates (Remmers & Elliott, 1949).]
3. Whether a course was a required course or an elective. [Teachers of required courses have been found to receive lower ratings than teachers of electives (Gage, 1961; Lovell & Haner, 1955).]
4. Whether a course belonged to the major area of study of the rater. [There was a tendency for majors to give more favorable ratings than non-majors (Cohen & Humphreys, 1960).]

Another set of variables have shown little relationship to an instructor's evaluation: (a) the difficulty of the course (Remmers, 1928); (b) the sex of student raters and of instructors (Bendig, 1953b; Caffrey, 1969; Downie, 1952; Elliott, 1950; Heilman & Armentrout, 1936; Lovell & Haner, 1955; Remmers, 1939); and (c) the popularity in extraclass activities of the teacher (Remmers,

1928; 1960). The teacher's research productivity and his judged teaching effectiveness have not been found to be related (Aleamoni & Yimer, 1972; Guthrie, 1949, 1954; Hayes, 1971; Voeks, 1962), although the relationship may vary if authorship of books or that of articles is considered (McDaniel & Feldhusen, 1970). The relationship between size of class and rating is not clear. Some researchers (e.g., Heilman & Armentrout, 1936; Lovell and Haner, 1955) have provided support for this relationship while others such as Goodhartz (1948), Graham (1972), and Guthrie (1954) have failed to find evidence for it.

One variable which has not been examined is the information available to students concerning the purpose of the rating when they are requested to complete a course evaluation form. Such information may be provided by the administrator of the questionnaire or may be the result of hearsay or comments made by fellow students at some time in the past. A questionnaire designed to elicit opinions about student ratings of instructors was administered to 404 students enrolled in psychology courses at the University of Illinois. A substantial percentage (45%) of the students believed that the results of ratings would affect a teacher's future classroom behavior but many (59%) doubted that it would affect the teacher's status or advancement (Costin, Greenough, & Menges, 1971). An important issue raised by these findings is whether student ratings of instructors are affected by what is perceived to be the ultimate use of student ratings. The present study addresses itself to this question: Would students who were told that the results of the evaluation were to be used to determine the salary and the promotion of the instructor respond differently on a course evaluation form than students who were not given this information? It was hypothesized that students who were provided with such information would rate the instructor more favorably than students lacking this information.

Method

Subjects

The subjects who participated in the study were two groups of undergraduate students who were enrolled in a history course, taught by the same instructor, during the fall 1971 semester and the spring 1972 semester and who completed a course evaluation form. One hundred and sixty-four of the 277 students who were enrolled in the fall semester and 80 of the 193 students who were registered in the spring semester completed the questionnaire.

The percentage composition of each of the classes, based on those who completed the evaluation, is as follows:

	Freshman	Sophomore	Junior	Senior	Graduate
Fall	29	21	09	02	00
Spring	38	32	14	15	00

Thirty-eight percent of the fall students and one percent of the spring students omitted the question on status. One percent of the fall students marked the "other" category.

The corresponding composition for the entire group of students registered for the course is presented below:

	Freshman	Sophomore	Junior	Senior	Graduate
Fall	43	31	15	11	00
Spring	34	30	18	18	00

A χ^2 test to compare sample and population distributions was performed separately for each semester (Only the categories of Freshman, Sophomore, Junior and Senior were considered.). Neither for the fall ($\chi^2 = 6.68$, $v = 3$, $p > .05$) nor for the spring ($\chi^2 = 1.43$, $v = 3$, $p > .50$) were there any significant

differences in composition between those students who completed the questionnaire and the whole group of students who were registered for the course. Thus, in terms of the level of study, each sample that evaluated the course is considered representative of the entire group of students registered in the course for that particular semester.

Materials

The questionnaire that was used to evaluate the course was the Illinois Course Evaluation Questionnaire (CEQ). The CEQ was developed to "elicit student opinions about a standardized set of statements relative to certain standardized aspects of an instructional program" (Spencer & Aleamoni, 1969). The CEQ consists of fifty items. The reliability of the total test has been calculated as .93 using a Spearman-Brown correlation corrected for length (Spencer & Aleamoni, 1970) and .98 using Cronbach's α on more recent data (Gillmore, 1973). The fifty items of the CEQ are grouped into six subscores. Five of the subscores were developed by factor analysis and the sixth consists of items that did not load highly on the other factors but were retained because of their special interest to faculty memebers. The six subscores are presented in Table 1.

Table 1

Definition of the Six Subscores

Number	Subscore	No. of Items
1	General Course Attitude	8
2	Method of Instruction	8
3	Course Content	8
4	Interest and Attention	8
5	Instructor	8
6	Specific Items	10

The product-moment correlations between the subscores usually range from .46 to .77. The CEQ is printed on machine-scorable answer sheets. There are four response positions for each question: Strongly agree, agree, disagree, and strongly disagree. The items are either stated negatively or positively. For positive statements a weight of 4, 3, 2, and 1 is assigned, respectively, for the response position, while for negatively worded statements, the reverse is true.

Procedure

The CEQ was administered to each group of students during the last week of each semester. The final examination for the course had not yet taken place. In the fall the instructor read to the students a set of standard instructions for marking the CEQ forms. These instructions pertain to procedural aspects of filling in the questionnaire (e.g., inserting the course code number and using pencil only) but do not include a description of the CEQ or its purpose. Before administering the questionnaire in the spring, the instructor provided the students with information about the CEQ. Its purpose was described as "to help your instructor, department, or college determine your attitudes, interests, and opinions related to this course and the way the course material was presented." "...through procedures such as these, improvements in instruction, course material and methods of teaching may occur." Moreover, the students were told that the results would be used "...for salary and promotion consideration of your instructor by his department head." The set of standard instructions followed. For both administrations of the questionnaire, the students were instructed <u>not</u> to mark their student number, in order to assure anonymity.

Results

Before analyzing the students' responses on the CEQ, a comparison was made of the two groups of students, to determine if they differed from each other. The students were compared on seven measures--four ACT (American College Testing Program) scores, two SCAT (School and College Ability Tests) scores, and High School Percentile Rank. The ACT battery consists of four examinations--English Usage, Mathematics Usage, Social Sciences Reading, and Natural Sciences Reading. The aim of the tests is to measure as directly as possible the abilities that the student will need to apply in his college course work. The SCAT tests measure verbal ability, or understanding of words, and mathematical ability, or understanding of fundamental number operations. ACT scores, SCAT scores, and High School Percentile Rank are used to predict success in college.

Complete sets of scores were obtained for 241 of the 277 students enrolled in the fall semester and for 163 of the 193 students enrolled in the spring semester. A discriminant analysis performed on the seven scores revealed that the difference between the two groups just reached significance ($F = 2.04$, $df = 7,396$, $p = .05$). The discriminant function is presented below:

$$V \text{ (normalized)} = .38084x_1 - .17698x_2 + .36721x_3 - .80509x_4 + .19536x_5 - .02973x_6 + .03970x_7$$

$$V \text{ (standardized)} = 30.924x_1 - 20.519x_2 + 36.427x_3 - 81.394x_4 + 34.346x_5 - 5.7025x_6 + 12.525x_7$$

where the x's in order of occurrence are: ACT English Usage, ACT Mathematics Usage, ACT Social Sciences Reading, ACT Natural Sciences Reading, SCAT Verbal

Ability, SCAT Quantitiative Ability, and High School Percentile Rank. The means for the two groups on the discriminant function are -.13 for the spring and 1.28 for the fall.

As can be seen, the largest single weight contributing to the discriminant function is that of the Natural Science score. A univariate t test for large samples performed on the Natural Science score showed that the two groups did not differ significantly on this measure ($t = 1.71$, $df = 402$, $p > .05$). Similarly, univariate t tests on the other measures failed to find significant differences between the two groups. The t values for the first three measures and the last three are, respectively, $t = .97$, $t = -1.10$, $t = .63$, $t = 1.34$, $t = -1.04$, $t = .84$. All occur with a probability greater than .05.

Next, because of the possibility that actual or expected grades might influence the rating of the course, an examination of the grades was necessary. A comparison was undertaken of the grades that the students expected to receive and the grades that were actually received, for both semesters. Neither for the spring ($t = 1.87$, $df = 261$, $p > .05$) nor for the fall ($t = .33$, $df = 375$, $p > .05$) semester were there any significant differences between actual and expected grades. Also, there were no significant differences between the mean actual grade for the fall and the mean actual grade for the spring ($t = 1.15$, $df = 458$, $p > .05$). However, the difference between the mean expected grade of the fall and of the spring was significant ($t = 2.40$, $df = 178$, $p < .01$). Students enrolled in the spring expected to receive higher grades than those enrolled in the fall.

Effects of Differential Instruction

To assess the effects of the differential instructions, six subscale scores were computed for each individual (see Table 1). These were the

dependent measures. A discriminant analysis was performed on the subscale scores for the fall and spring semesters. The following discriminant function was obtained:

$$V \text{ (normalized)} = .32779x_1 + .37955x_2 + .13438x_3 - .26797x_4 + .41880x_5 + .69515x_6$$

$$V \text{ (standardized)} = 3.2459x_1 + 3.6416x_2 + .89542x_3 - 2.8117x_4 + 2.8176x_5 + 3.9003x_6$$

The group means on the discriminant function are 4.62 for the fall and 5.01 for the spring. The overall F was significant ($F = 3.47$, $df = 6,237$, $p < .01$). Individual t tests for large samples were performed for each of the six separate subscale scores. All differences between the groups were significant at the .01 level. The respective values of t are: $t = 4.03$, $t = 3.99$, $t = 3.31$, $t = 3.49$, $t = 3.55$, $t = 4.06$.

Discussion

As was hypothesized, the students enrolled in the spring semester who were informed that the course evaluation would be used for salary and promotion consideration of the instructor, rated the course more favorably than students enrolled in the fall who were given the standard set of instructions. Using a discriminant analysis, differences between the two groups of students on ACT, SCAT and High School Percentile Rank just reached significance at the .05 level while individual t tests proved nonsignificant. Thus, differences between the groups are not believed to have contributed practically or statistically to the differential evaluation of the instructor and the course. Similarly, student evaluations were not influenced by actual grades received. Although students in the spring expected to receive higher grades than those who were registered in the fall, the absolute magnitude of difference (3.91 versus 3.64 on a 5 point scale) was very small and is

unlikely to have affected the rating of the instructor.

The importance of the finding that students who were informed of the use of the evaluation for salary and promotion purposes rated the instructor more highly is emphasized by the fact that the course evaluation was higher on each of the six subscores. When this course was compared with other courses at the same level, other courses at the same institution, and other courses across the United States that have employed the CEQ [For a description of the different norms, see the CEQ manual (Aleamoni, 1972b).], the spring evaluation for each of the subscales was seen to rise by an average of two deciles.

This study will need to be replicated with other levels of courses and other types of courses and with larger numbers of students before it will be possible to generalize about the effects of different sets of instructions on course evaluation. However, the results do suggest that the instructions and the information about the aims of a course evaluation that students are provided prior to the completion of the evaluation are important variables that will have to be considered in future administrations of any course evaluation.

References

Aleamoni, L. M. A review of some recent reliability and validity studies on the Illinois Course Evaluation Questionnaire (CEQ). Research Memorandum No. 127, Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1972. (a)

Aleamoni, L. M. Illinois Course Evaluation Questionnaire: Results interpretation manual, Form 66 and Form 32. Research Report No. 331, Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1972. (b)

Aleamoni, L. M. & Yimer, M. An investigation of the relationship between colleague rating, student rating, research productivity, and academic rank in rating instructional effectiveness. Research Report No. 338, Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1972.

Anikeeff, A. M. Factors affecting student evaluation of college faculty members. Journal of Applied Psychology, 1953, 37, 458-460.

Bendig, A. W. Relation of level of course achievement of students, instructor and course ratings in introductory psychology. Educational and Psychological Measurement, 1953, 13, 437-488. (a)

Bendig, A. W. Student achievement in introductory psychology and student ratings of the competence and empathy of their instructors. Journal of Psychology, 1953, 36, 427-433. (b)

Caffrey, B. Lack of bias in student evaluations of teachers. Proceedings of the 77th Annual Convention of the American Psychological Association, 1969, 4, 641-642.

Clark, K. E. & Keller, R. J. Student ratings of college teaching. In R. A. Eckert (Ed.), A university looks at its program. Minneapolis: University of Minnesota Press, 1954.

Cohen, J. & Humphreys, L. G. Memorandum to faculty. University of Illinois, Department of Psychology, 1960. (mimeographed)

Costin, F., Greenough, W. T. & Menges, R. J. Student ratings of college teaching: Reliability, validity, and usefulness. Review of Educational Research, 1971, 41, 511-535.

Downie, N. W. Student evaluation of faculty. Journal of Higher Education, 1952, 23, 495-496, 503.

Elliott, D. H. Characteristics and relationships of various criteria of college and university teaching. Purdue University Studies in Higher Education, 1950, 70, 5-61.

Gage, N. L. The appraisal of college teaching. Journal of Higher Education, 1961, 32, 17-22.

Gillmore, G. M. Estimates of reliability coefficients for items and subscales of the Illinois Course Evaluation Questionnaire. Research Report No. 341. Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1973.

Goodhartz, A. S. Student attitudes and opinions relating to teaching at Brooklyn College. School and Society, 1948, 68, 345-349.

Graham, M. H. The relationship between CEQ ratings and instructor's rank, class size, and course level. Research Report #337, Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1972.

Guthrie, E. R. The evaluation of teaching. Educational Record, 1949, 30, 109-115.

Guthrie, E. R. The evaluation of teaching: A progress report. Seattle: University of Washington, 1954.

Hayes, J. R. Research, teaching and faculty fate. Science, 1971, 172, 227-230.

Heilman, J. D. & Armentrout, W. D. The rating of college teachers on ten traits by their students. Journal of Educational Psychology, 1936, 27, 197-216.

Lovell, G. D. & Haner, C. F. Forced-choice applied to college faculty rating. Educational and Psychological Measurement, 1955, 15, 291-304.

McDaniel, E. D. & Feldhusen, J. F. Relationships between faculty ratings and indexes of service and scholarship. Proceedings of the 78th Annual Convention of the American Psychological Association, 1970, 5, 619-620.

Rayder, N. F. College student ratings of instructors. Journal of Experimental Education, 1968, 37, 76-81.

Remmers, H. H. The relationship between students' marks and students' attitudes toward instructors. School and Society, 1928, 28, 759-760.

Remmers, H. H. To what extent do grades influence student ratings of instructors? Journal of Educational Research, 1930, 21, 314-316.

Remmers, H. H. Appraisal of college teaching through ratings and student opinion. In 27th Yearbook of the National Society of College Teachers of Education. Chicago: University of Chicago Press, 1939.

Remmers, H. H. Manual of instructions for the Purdue Rating Scale for Instructors (Rev. ed.). West Lafayette, Indiana: University Book Store, 1960.

Remmers, H. H. & Elliott, D. N. The Indiana college and university staff-evaluation program. School and Society, 1949, 70, 168-171.

Spencer, R. E. & Aleamoni, L. M. The Illinois Course Evaluation Questionnaire: A description of its development and a report of some of its results. Research Report No. 292. Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1969.

Spencer, R. E. & Aleamoni, L. M. A student course evaluation questionnaire. Journal of Educational Measurement, 1970, 7, 209-210.

Stewart, C. T. & Malpass, L. F. Estimates of achievement and ratings of instructors. Journal of Educational Research, 1966, 59, 347-350.

Voeks, V. W. Publications and teaching effectiveness. Journal of Higher Education, 1962, 33, 212.

Weaver, C. H. Instructor rating by college students. Journal of Educational Psychology, 1960, 51, 21-25.